UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
Department of Electrical and Computer Engineering

ECE 417 Multimedia Signal Processing
Spring 2021

PRACTICE EXAM 2

Exam will be Tuesday, November 2, 2021

- This will be a CLOSED BOOK exam.
- You will be permitted one sheet of handwritten notes, 8.5x11.
- Calculators and computers are not permitted.
- If you're taking the exam online, you will need to have your webcam turned on. Your exam will appear on Gradescope at exactly 9:30am; you will need to photograph and upload your answers by exactly 11:00am.
- There will be a total of 100 points in the exam. Each problem specifies its point total. Plan your work accordingly.
- You must SHOW YOUR WORK to get full credit.

1. (16 points) Suppose you have an $M \times D$ matrix, $X = [\vec{x}_0, \ldots, \vec{x}_{M-1}]^T$, where $\sum_{m=0}^{M-1} \vec{x}_m = \vec{0}$. The eigenvalues of $X^T X$ are λ_0 through λ_{D-1}, its eigenvectors are $\vec{v}_0$ through $\vec{v}_{D-1}$, and its principal components are $Y = XV$.

   (a) Write $Y^T Y$ in terms of the eigenvalues, λ_0 through λ_{D-1}.

   Solution:

$$Y^T Y = \begin{bmatrix} \lambda_0 & 0 & \ldots & 0 \\ 0 & \lambda_1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \lambda_{D-1} \end{bmatrix}$$

   (b) Write $\sum_{m=0}^{M-1} \|\vec{x}_m\|_2^2$ in terms of the eigenvalues, λ_0 through λ_{D-1}.

   Solution:

$$\sum_{m=0}^{M-1} \|\vec{x}_m\|_2^2 = \sum_{d=0}^{D-1} \lambda_d$$

   (c) Write $\vec{v}_i^T X^T X \vec{v}_j$ in terms of the eigenvalues, λ_0 through λ_{D-1}, for $0 \leq i \leq j \leq D-1$.

   Solution:

$$\vec{v}_i^T X^T X \vec{v}_j = \begin{cases} \lambda_i & i = j \\ 0 & \text{otherwise} \end{cases}$$

2. (16 points) A 2-dimensional Gaussian random vector has mean $\vec{\mu}$ and covariance Σ given by

$$\vec{\mu} = \left[\begin{array}{c} 1 \\ 1 \end{array}\right], \quad \Sigma = \left[\begin{array}{cc} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{array}\right] \left[\begin{array}{cc} 8 & 0 \\ 0 & 2 \end{array}\right] \left[\begin{array}{cc} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \end{array}\right]$$

Draw a curve of some kind, on a two-dimensional Cartesian plane, showing the set of points $\left\{\vec{x} : p_X(\vec{x}) = \frac{1}{8\pi} e^{-\frac{1}{2}}\right\}$.

Solution: $|\Sigma| = |\Lambda| = 16$, so

$$p_X(\vec{x}) = \frac{1}{8\pi} e^{-\frac{1}{2} d_\Sigma^2(\vec{x},\vec{\mu})}$$

so the solution is the set $\left\{\vec{x} : d_\Sigma^2(\vec{x},\vec{\mu}) = 1\right\}$.

$$d_\Sigma^2(\vec{x},\vec{\mu}) = \vec{y}^T \Lambda^{-1} \vec{y} = \frac{y_1^2}{8} + \frac{y_2^2}{2}$$

$$\vec{y} = \frac{\sqrt{2}}{2} \left[\begin{array}{c} (x_1 - 1) + (x_2 - 1) \\ (x_1 - 1) - (x_2 - 1) \end{array}\right]$$

So the solution is the set

$$\left\{\vec{x} : \frac{(x_1 + x_2 - 2)^2}{16} + \frac{(x_1 - x_2)^2}{4} = 1\right\}$$

...which is an ellipse, centered at $(1,1)$, with a radius of $2\sqrt{2}$ along the $(1,1)$ direction, and a radius of $\sqrt{2}$ along the $(1,-1)$ direction.

3. (16 points) A particular HMM-based speech recognizer only knows two words: word w_0, and word w_1. Word w_0 has a higher *a priori* probability: $p_Y(w_0) = 0.7$, while $p_Y(w_1) = 0.3$. Each of the two words is modeled by a four-state Gaussian HMM ($N = 4$) with three-dimensional observations ($D = 3$). All states, in both HMMs, have identity covariance ($\Sigma_i = I$). Both HMMs have *exactly* the same transition probabilities and state-dependent means, given by:

$$\textbf{Both Words: } A = \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix}, \quad \vec{\mu}_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{\mu}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{\mu}_3 = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}, \quad \vec{\mu}_4 = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

But the initial residence probabilities are different:

$$\textbf{Word 0: } \pi_i = \begin{cases} 1 & i = 1 \\ 0 & \text{otherwise} \end{cases} \qquad \textbf{Word 1: } \pi_i = \begin{cases} 1 & i = 4 \\ 0 & \text{otherwise} \end{cases}$$

Suppose that you have a two-frame observation, $X = [\vec{x}_1, \vec{x}_2]$, where $\vec{x}_t = [x_{1t}, x_{2t}, x_{3t}^T]$. The MAP decision rule, in this case, can be written as a linear classifier,

$$\hat{y} = \begin{cases} w_1 & \vec{w}_1^T \vec{x}_1 + \vec{w}_2^T \vec{x}_2 + b > 0 \\ w_0 & \text{otherwise} \end{cases}$$

Find $\vec{w}_1$, $\vec{w}_2$, and b.

Solution: The Bayesian classifier chooses w_1 if

$$\begin{aligned} p(w_0)p(X|w_0) &< p(w_1)p(X|w_1) \\ 0.7\mathcal{N}(\vec{x}_1|\vec{\mu}_1)\sum_j a_{1j}\mathcal{N}(\vec{x}_2|\vec{\mu}_j) &< 0.3\mathcal{N}(\vec{x}_4|\vec{\mu}_4)\sum_j a_{4j}\mathcal{N}(\vec{x}_2|\vec{\mu}_j) \\ 0.7\mathcal{N}(\vec{x}_1|\vec{\mu}_1) &< 0.3\mathcal{N}(\vec{x}_1|\vec{\mu}_4) \\ \ln(0.7) - \frac{1}{2}(\vec{x}_1 - \vec{\mu}_1)^T(\vec{x}_1 - \vec{\mu}_1) &< \ln(0.3) - \frac{1}{2}(\vec{x}_1 - \vec{\mu}_4)^T(\vec{x}_1 - \vec{\mu}_4) \\ \ln(0.7) - \frac{1}{2}\|\vec{x}_1\|^2 &< \ln(0.3) - \frac{1}{2}\|\vec{x}_1\|^2 + \vec{\mu}_4^T\vec{x}_1 - \frac{1}{2}\|\vec{\mu}_4\|^2 \end{aligned}$$

Which is satisfied if

$$\vec{\mu}_4^T\vec{x}_1 + \ln\left(\frac{3}{7}\right) - \frac{3}{2} > 0$$

So

$$\vec{w}_1 = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}, \quad \vec{w}_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad b = \ln\left(\frac{3}{7}\right) - \frac{3}{2}$$

4. (16 points) In terms of $\alpha_t(i)$, $\beta_t(i)$, a_{ij}, π_i and $b_i(\vec{x}_t)$, find

$$p\left(q_6 = i, q_7 = j | \vec{x}_1, \ldots, \vec{x}_{20}\right)$$

Solution:

$$\begin{aligned} p(q_6 = i, q_7 = j, \vec{x}_1, \ldots, \vec{x}_{20}) &= p(\vec{x}_1, \ldots, \vec{x}_6, q_6 = i) p(q_7 = j | q_6 = i) p(\vec{x}_7 | q_7 = j) p(\vec{x}_8, \ldots, \vec{x}_{20} | q_7 = j) \\ &= \alpha_6(i) a_{ij} b_j(\vec{x}_7) \beta_7(j) \end{aligned}$$

$$p\left(q_6 = i, q_7 = j | \vec{x}_1, \ldots, \vec{x}_{20}\right) = \frac{\alpha_6(i) a_{ij} b_j(\vec{x}_7) \beta_7(j)}{\sum_{k=1}^{N} \sum_{\ell=1}^{N} \alpha_6(k) a_{k\ell} b_\ell(\vec{x}_7) \beta_7(\ell)}$$

5. (5 points) Suppose you have a dataset including the vectors

$$\vec{x} = \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}, \quad \vec{y} = \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix}, \quad \vec{z} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

Find a diagonal matrix Σ such that $d_\Sigma^2(\vec{x}, \vec{y}) > d_\Sigma^2(\vec{x}, \vec{z})$.

Solution: Any solution such that $\frac{1}{\sigma_1^2} > \frac{1}{\sigma_2^2} + \frac{1}{\sigma_3^2}$

6. (10 points) Define $\Phi(z)$ as follows:

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$$

Suppose $\vec{X} = [X_1, X_2]^T$ is a Gaussian random vector with mean and covariance given by

$$\vec{\mu} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 9 & 0 \\ 0 & 4 \end{bmatrix}$$

(a) Sketch the set of points such that $f_{\vec{X}}(\vec{x}) = \frac{1}{12\pi} e^{-\frac{1}{8}}$, where $f_{\vec{X}}(\vec{x})$ is the pdf of $\vec{X}$.

Solution: The sketch should show an ellipse with axes parallel to the main axes, passing through the points $(\frac{5}{2}, 0)$, $(-\frac{1}{2}, 0)$, $(1, 1)$, and $(1, -1)$.

(b) In terms of $\Phi(z)$, find the probability $\Pr\{-1 < X_1 < 1, -1 < X_2 < 1\}$.

Solution: $(\Phi(0) - \Phi(-\frac{2}{3}))(\Phi(\frac{1}{2}) - \Phi(-\frac{1}{2}))$

7. (10 points) Suppose that a particular covariance matrix Σ has the following eigenvector matrix, U, and eigenvalue matrix, Λ:

$$U = \frac{\sqrt{2}}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad \Lambda = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$$

Let $\vec{y}(\vec{x}) = \begin{bmatrix} y_1(\vec{x}) \\ y_2(\vec{x}) \end{bmatrix} = U^T\vec{x}$ be the principal components of a vector space $\vec{x}$.

(a) Plot the set of vectors $\vec{x}$ such that $y_1(\vec{x}) = 3$.

Solution: The sketch should show the line $x_1 + x_2 = 3\sqrt{2}$.

(b) Find the squared Mahalanobis distance, $d^2_\Sigma(\vec{x}, \vec{\mu})$, between the vectors $\vec{x}$ and $\vec{\mu}$ where

$$\vec{x} = \begin{bmatrix} 5 \\ 5 \end{bmatrix}, \quad \vec{\mu} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Solution: 8

8. (10 points) Suppose that, for a particular classification problem, the correct label of every data point is as follows:

$$y^*(\vec{x}) = \begin{cases} 1 & \|\vec{x}\|_2 < 1.5 \\ 0 & \|\vec{x}\|_2 > 1.5 \end{cases} \tag{1}$$

Unfortunately, you aren't allowed to use the correct labeling function. Instead, you have to try to learn a nearest-neighbor or Bayesian classifier.

(a) Your nearest-neighbor classifier is trained using 25 training samples, taken at integer coordinates for $-2 \le x_1, x_2 \le 2$. Fortunately, your training data are correctly labeled, using the labeling function shown in Eq. (1). Thus the complete training dataset is

$$X = \left[\begin{array}{cccccccc} -2 & -2 & \dots & 0 & 0 & 0 & \dots & 2 \\ -2 & -1 & \dots & 0 & 1 & 2 & \dots & 2 \end{array}\right], \quad Y = [0, 0, \dots, 1, 1, 0, \dots, 0]$$

Using these 25 training examples, you construct a nearest-neighbor classifier. Draw the decision boundary of the resulting nearest-neighbor classifier.

Solution: The sketch should show the square $\max(|x_1|, |x_2|) = 1.5$.

(b) Suppose now that $f_{\vec{X}|Y}(\vec{x}|0)$ and $f_{\vec{X}|Y}(\vec{x}|1)$ are both zero-mean Gaussian pdfs, with the covariance matrices Σ_0 and Σ_1 respectfully, where

$$\Sigma_0 = \left[\begin{array}{cc} 2 & 0 \\ 0 & 2 \end{array}\right], \quad \Sigma_1 = \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right]$$

Define η to be the odds ratio, $\eta = p_Y(0)/p_Y(1)$. Find a value of η such that a Bayesian classifier gives exactly the decision boundary shown in Eq. (1) on the previous page.

Solution: $y^* = 1$ if

$$\begin{aligned} \ln p_Y(1) - \frac{1}{2}\ln|\Sigma_1| - \frac{1}{2}\vec{x}^T\Sigma_1^{-1}\vec{x} &> \ln p_Y(0) - \frac{1}{2}\ln|\Sigma_0| - \frac{1}{2}\vec{x}^T\Sigma_0^{-1}\vec{x} \\ \ln p_Y(1) - \frac{1}{2}\ln(1) - \frac{1}{2}\vec{x}^T\vec{x} &> \ln p_Y(0) - \frac{1}{2}\ln(4) - \frac{1}{4}\vec{x}^T\vec{x} \\ 0 &> \ln\eta - \frac{1}{2}\ln(4) + \frac{1}{4}\vec{x}^T\vec{x} \\ 4\ln(2/\eta) &> \|\vec{x}\|_2^2 \end{aligned}$$

This matches the criterion $\|\vec{x}\|_2 < 1.5$ if

$$\begin{aligned} 4\ln(2/\eta) &= (1.5)^2 = \frac{9}{4} \\ \ln(2/\eta) &= \frac{9}{16} \\ \eta &= 2e^{-9/16} \end{aligned}$$

9. (10 points) Suppose that you have M different D-dimensional vectorized face images, $\vec{\Gamma}_m = [\gamma_{1m}, \ldots, \gamma_{Dm}]^T$, whose mean is $\vec{\Psi} = [\psi_1, \ldots, \psi_D]^T$. Define the data matrix to be $A = [\vec{\Gamma}_1 - \vec{\Psi}, \ldots, \vec{\Gamma}_M - \vec{\Psi}]$, and suppose that the eigenvectors and eigenvalues of $A^T A$ are given by $U = [\vec{u}_1, \ldots, \vec{u}_M]$ and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_M)$.

(a) Find the numerical value of the vector $U^T \vec{u}_3$.

Solution: $U^T \vec{u}_3 = [0, 0, 1, 0, \ldots, 0]^T$

(b) Your goal is to find a $(D \times M)$ matrix $V = [\vec{v}_1, \ldots, \vec{v}_M]$ so that $\vec{\Omega}_m = V^T(\vec{\Gamma}_m - \vec{\Psi})$ is a vector containing the first M principal components of the image $\vec{\Gamma}_m$. Write an equation showing how V can be computed from $\vec{\Psi}$, A, U, and/or Λ.

Solution: $V \propto AU$ (any constant of proportionality is an acceptable answer).

10. (10 points) Suppose that you have M different D-dimensional vectorized face images, $\vec{\Gamma}_m = [\gamma_{1m}, \ldots, \gamma_{Dm}]^T$, whose mean is $\vec{\Psi} = [\psi_1, \ldots, \psi_D]^T$. Define the scatter matrix to be

$$S = \sum_{m=1}^{M} (\vec{\Gamma}_m - \vec{\Psi})(\vec{\Gamma}_m - \vec{\Psi})^T$$

Suppose that the eigenvectors and eigenvalues of S are $V = [\vec{v}_1, \ldots, \vec{v}_D]$ and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_D)$. You want to find a value of K such that the K-dimensional PCA projection $\vec{\Omega}_m = [\vec{v}_1, \ldots, \vec{v}_K]^T(\vec{\Gamma}_m - \vec{\Psi})$ has the following property:

$$\sum_{m=1}^{M} \left|\vec{\Omega}_m\right|^2 = (0.95) \sum_{m=1}^{M} \left|\vec{\Gamma}_m - \vec{\Psi}\right|^2 \tag{2}$$

Specify an equation that, if satisfied, will guarantee the truth of Eq. 2. Your equation should only include the scalars M, D, K, and/or the eigenvalues λ_d ($1 \leq d \leq D$); your equation should not include $\vec{\Gamma}_m$ or $\vec{\Psi}$.

Solution:

$$\sum_{k=1}^{K} \lambda_k = (0.95) \sum_{k=1}^{M} \lambda_k$$

11. (20 points) A particular dataset has three data,

$$\vec{x}_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \quad \vec{x}_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \quad \vec{x}_3 = \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Define $X = [\vec{x}_1, \vec{x}_2, \vec{x}_3]$ and $R = X^T X$. The matrix R is given by $R = V \Lambda V^T$ where

$$V = \begin{bmatrix} -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{2}{\sqrt{6}} & 0 \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \end{bmatrix}, \quad \Lambda = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$$

Find a matrix W such that $\vec{y}_i = W^T \vec{x}_i$, $\vec{y}_i$ is two-dimensional, and the elements of $\vec{y}_i$ are uncorrelated.

Solution:

$$W = \begin{bmatrix} 0 & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \\ \sqrt{\frac{3}{2}} & -\frac{1}{\sqrt{2}} \\ 0 & 0 \\ -\sqrt{\frac{3}{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

...or any matrix whose columns are proportional to the columns shown above (because the problem asked for uncorrelated elements, but did not specify the variance of each element).

12. (16 points) A particular dataset has the scatter matrix $S = \sum_{k=1}^{n} (\vec{x}_k - \vec{m})(\vec{x}_k - \vec{m})^T$, whose first two eigenvectors are $\vec{v}_1$ and $\vec{v}_2$, characterized by eigenvalues $\lambda_1 = 450$ and $\lambda_2 = 150$. Define the transform $\vec{y}_k = [\vec{v}_1, \vec{v}_2]^T (\vec{x}_k - \vec{m})$. Define the 2×2 matrix

$$Q = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix} = \sum_{k=1}^{n} \vec{y}_k \vec{y}_k^T$$

Find the numerical values of the elements q_{11}, q_{12}, q_{21}, and q_{22} of matrix Q.

Solution:

$$\begin{aligned} Q &= \sum_{k=1}^{n} V^T (\vec{x}_k - \vec{m})(\vec{x}_k - \vec{m})^T V \\ &= V^T S V = \Lambda \\ &= \begin{bmatrix} 450 & 0 \\ 0 & 150 \end{bmatrix} \end{aligned}$$

13. (16 points) A particular dataset has six data vectors, given by

$$\{\vec{x}_1, \ldots, \vec{x}_6\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \right\}$$

By calling `np.random.randn`, you generate a 3×2 random projection matrix V, given by

$$V = \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \\ v_{31} & v_{32} \end{bmatrix}$$

Using this random projection matrix, you compute the transformed feature vectors $\vec{y}_k = V^T \vec{x}_k$. The total energy of the transformed dataset can be written as

$$E = \sum_{k=1}^{6} \vec{y}_k^T \vec{y}_k$$

Find the value of E in terms of the random projection matrix elements v_{ij}.

Solution:

$$Y = V^T X = \begin{bmatrix} v_{11} & -v_{11} & v_{21} & -v_{21} & v_{31} & -v_{31} \\ v_{12} & -v_{12} & v_{22} & -v_{22} & v_{32} & -v_{32} \end{bmatrix}$$

$$E = \sum_{k=1}^{6} \|\vec{y}_k\|^2 = 2 \sum_{i=1}^{3} \sum_{j=1}^{2} v_{ij}^2$$

14. (16 points) You want to classify zoo animals. Your zoo only has two species: elephants and giraffes. There are more elephants than giraffes: if Y is the species,

$$\begin{aligned} p_Y(\text{elephant}) &= \frac{e}{e+1} \\ p_Y(\text{giraffe}) &= \frac{1}{e+1} \end{aligned}$$

where $e = 2.718\ldots$ is the base of the natural logarithm. The height of giraffes is Gaussian, with mean $\mu_G = 5$ meters and variance $\sigma_G^2 = 1$. The height of elephants is also Gaussian, with mean $\mu_E = 3$ and variance $\sigma_E^2 = 1$. Under these circumstances, the minimum probability of error classifier is

$$\hat{y}(x) = \begin{cases} \text{giraffe} & x > \theta \\ \text{elephant} & x < \theta \end{cases}$$

Find the value of θ that minimizes the probability of error.

Solution: $\theta = 4.5$

15. (16 points) Random vector X is distributed as

$$p_X(\vec{x}) = \sum_{k=1}^{2} c_k \mathcal{N}(\vec{x}|\vec{\mu}, \Sigma)$$

where $c_1 = c_2 = 0.5$, and

$$\vec{\mu}_1 = \begin{bmatrix} -2 \\ 0 \end{bmatrix}, \quad \vec{\mu}_2 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \quad \Sigma_1 = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}, \quad \Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$$

Draw a contour plot showing $p_X(\vec{x})$ as a function of $\vec{x}$. Mark the modes of the distribution, and draw contour lines at levels of $e^{-1/2}$ and e^{-2} times the height of the modes.

Solution: Modes of the distribution are at $[-2, 0]^T$ and $[2, 0]^T$. The $e^{-1/2}$ contour lines are ellipses: a 4×2 ellipse centered at $[-2, 0]^T$, and a 2×4 ellipse centered at $[2, 0]^T$. The e^{-2} contour line is the continuous outer hull of the 8×4 and 4×8 ellipses centered at the modes.

16. (16 points) A particular hidden Markov model is parameterized by $\lambda = \{\pi_i, a_{ij}, b_j(\vec{x})\}$ where π_i is uniform ($\pi_i = \frac{1}{N}$). Devise an algorithm to compute $p(q_1 = k|\vec{x}_1, \ldots, \vec{x}_T, \lambda)$. Your algorithm should be similar to the forward algorithm, but with a different initialization.

Solution: There are several possible solutions. One is

$$\begin{aligned}
p(q_1 = k, \vec{x}_1|\lambda) &= \frac{1}{N} b_k(\vec{x}_1) \\
p(q_2 = i, q_1 = k, \vec{x}_1, \vec{x}_2|\lambda) &= p(q_1 = k, \vec{x}_1|\lambda) a_{ki} b_i(\vec{x}_2) \\
p(q_t = j, q_1 = k, \vec{x}_1, \ldots, \vec{x}_t|\lambda) &= \sum_{i=1}^{N} p(q_{t-1} = i, q_1 = k, \vec{x}_1, \ldots, \vec{x}_{t-1}|\lambda) a_{ij} b_j(\vec{x}_t) \\
p(q_1 = k, \vec{x}_1, \ldots, \vec{x}_T|\lambda) &= \sum_{j=1}^{N} p(q_T = j, q_1 = k, \vec{x}_1, \ldots, \vec{x}_T|\lambda)
\end{aligned}$$

$$p(q_1 = k|\vec{x}_1, \ldots, \vec{x}_T, \lambda) = \frac{p(q_1 = k, \vec{x}_1, \ldots, \vec{x}_T|\lambda)}{\sum_{\ell=1}^{N} p(q_1 = \ell, \vec{x}_1, \ldots, \vec{x}_T|\lambda)}$$

17. (16 points) The scaled forward algorithm is provided for you on the formula page at the beginning of this exam. In terms of the quantities $\pi_i, a_{ij}, b_j(\vec{x}), \hat{\alpha}_t(j), g_t$, and/or $\tilde{\alpha}_t(j)$, find a formula for the quantity $p(q_{t-1} = i, q_t = j, \vec{x}_{t-1}, \vec{x}_t | \vec{x}_1, \ldots, \vec{x}_{t-2}, \lambda)$.

Solution:

$$p(q_{t-1} = i, q_t = j, \vec{x}_{t-1}, \vec{x}_t | \vec{x}_1, \ldots, \vec{x}_{t-2}, \lambda) = \sum_{k=1}^{N} \hat{\alpha}_{t-2}(k) a_{ki} b_i(\vec{x}_{t-1}) a_{ij} b_j(\vec{x}_t)$$

18. (16 points) The stock market alternates between long bull markets (state 1) and short bear markets (state 2). This HMM has the following parameters:

$$\vec{\pi} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad A = \begin{bmatrix} 0.999 & 0.001 \\ 0.005 & 0.995 \end{bmatrix}, \quad \mu_1 = 0.1, \quad \mu_2 = -0.3, \quad \sigma_1^2 = \sigma_2^2 = 1,$$

where $\pi_i = p(q_1 = i)$, $a_{ij} = p(q_{t+1} = j | q_t = i)$, and $p(x_t | q_t = j) = \mathcal{N}(x_t; \mu_j, \sigma_j^2)$.

You observe x_2 on day 2.

For what values of x_2 does the forward algorithm yield probabilities $\alpha_t(i)$ such that $\alpha_2(2) > \alpha_2(1)$?

Asking exactly the same question in different words: for what values of x_2 would it be rational to conclude that a bear market has started?

Solution: $x_2 < -0.1 - 2.5 \ln(999)$

19. (16 points) Consider two PDFs. Class $y = 0$ is Gaussian:

$$p(x|y=0) = \mathcal{N}\left(x; \mu_0, \sigma_0^2\right)$$

Class $y = 1$ is mixture Gaussian, and for some reason, one of its mixture components is the Gaussian from class 0:

$$p(x|y=1) = 0.9p(x|y=0) + 0.1\mathcal{N}\left(x; \mu_1, \sigma_1^2\right)$$

where $\mu_0 = 0$, $\mu_1 = 3$, and $\sigma_0^2 = \sigma_1^2 = 1$.

For what values of x is

$$\frac{p(x|y=1)}{p(x|y=0)} > 1?$$

Solution: $x > \frac{3}{2}$

20. (16 points) A pelican fishes by sweeping its beak through the water. Each sweep catches many fish. The total weight of fish caught in a single sweep is an instance of a random variable, X, that is well described by a Gaussian mixture model:

$$p_X(x) = \sum_{k=1}^{2} c_k \mathcal{N}(x; \mu_k, \sigma_k^2)$$

Unfortunately, you don't know what are the correct values of the parameters c_k, μ_k, and σ_k.

(a) You have received the following suggestions for the parameters. For each candidate set of parameters, say whether or not $p_X(x)$ would be a valid probability density if computed using this set of parameters; if not, say why not.

i. Alice suggests $c_1 = 1, c_2 = 1, \mu_1 = 10, \mu_2 = 20, \sigma_1 = 10, \sigma_2 = 10$. Would $p_X(x)$ computed using this parameter set be a valid probability density? If not, why not?

> **Solution:** No, because $c_1 + c_2 \neq 1$.

ii. Barb suggests $c_1 = 0.1, c_2 = 0.9, \mu_1 = 0, \mu_2 = 20, \sigma_1 = 10, \sigma_2 = 10$. Would $p_X(x)$ computed using this parameter set be a valid probability density? If not, why not?

> **Solution:** yes.

iii. Carol suggests $c_1 = 0.5, c_2 = 0.5, \mu_1 = 10, \mu_2 = 20, \sigma_1 = -10, \sigma_2 = 10$. Would $p_X(x)$ computed using this parameter set be a valid probability density? If not, why not?

> **Solution:** Either no or yes is an acceptable answer, depending on your justification. If you said "no, because standard deviation can not be negative," that would be an acceptable answer. The correct answer, though, is actually "yes, because σ_1^2 is still positive, therefore a normal distribution computed using σ_1^2 as the variance would still be a valid pdf."

(b) You follow a pelican named Pete, and measure the weight of fish he retrieves on four consecutive dips, resulting in the following training dataset:

$$\{x_1, \ldots, x_4\} = \{5, 25, 15, 10\}$$

Using the parameter set $c_1 = 0.5, c_2 = 0.5, \mu_1 = 10, \mu_2 = 20, \sigma_1 = 10, \sigma_2 = 10$, compute $\gamma_k(x_t) = \Pr\left\{k^{\text{th}} \text{ Gaussian}|x_t\right\}$ for $1 \le t \le 4$, $1 \le k \le 2$. You might find the following table of Gaussian PDFs to be useful:

x	$\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$
0	0.40
0.5	0.35
1	0.24
1.5	0.13
2	0.05
2.5	0.02
3	0.00

Solution:

$$\begin{aligned}
\gamma_1(x_1) &= \frac{0.035}{0.035 + 0.013} \\
\gamma_1(x_2) &= \frac{0.013}{0.035 + 0.013} \\
\gamma_1(x_3) &= \frac{1}{2} \\
\gamma_1(x_4) &= \frac{0.04}{0.064} \\
\gamma_2(x_1) &= \frac{0.013}{0.035 + 0.013} \\
\gamma_2(x_2) &= \frac{0.035}{0.035 + 0.013} \\
\gamma_2(x_3) &= \frac{1}{2} \\
\gamma_2(x_4) &= \frac{0.024}{0.064}
\end{aligned}$$

(c) Recall that the training data are

$$\{x_1, \ldots, x_4\} = \{5, 25, 15, 10\}$$

Suppose that, after a few iterations of EM, you wind up with the following gamma probabilities:

$$\{\gamma_2(x_1), \gamma_2(x_2), \gamma_2(x_3), \gamma_2(x_4)\} = \{0.1, 0.8, 0.6, 0.6\}$$

Find the re-estimated values of c_2, μ_2, and σ_2^2 resulting from this iteration of EM.

Solution:

$$c_2 = \frac{2.1}{4}, \quad \mu_2 = \frac{(0.1)(5) + (0.8)(25) + (0.6)(15) + (0.6)(10)}{2.1},$$

$$\sigma_2^2 = \frac{(0.1)(5-\mu_2)^2 + (0.8)(25-\mu_2)^2 + 0.6(15-\mu_2)^2 + 0.6(10-\mu_2)^2}{2.1}$$

21. (16 points) You're training an audiovisual bird classifier: based on measurements of the birdsong frequency (f) and the bird color (c), the bird is classified as a sparrow ($s = 1$) if and only if

$$\eta \ln p(c|s=1) + (1-\eta) \ln p(f|s=1) > \eta \ln p(c|s=0) + (1-\eta) \ln p(f|s=0)$$

In truth, all sparrows have pitch $f < 0.5$, and color $c < 0.5$, while all other birds have pitch $f > 0.5$ and color $c > 0.5$. Unfortunately, your training algorithm is broken, so it learned these distributions:

$$p(f|s=0) = \begin{cases} 1 & 0 \leq f \leq 1 \\ 0 & \text{else} \end{cases}, \quad p(f|s=1) = \begin{cases} 1 & 0 \leq f \leq 1 \\ 0 & \text{else} \end{cases}, \quad p(c|s=0) = \begin{cases} 1 & 0 \leq c \leq 1 \\ 0 & \text{else} \end{cases}$$

In fact, only one of the pdfs was learned to be non-uniform:

$$p(c|s=1) = \begin{cases} 2-2c & 0 \leq c \leq 1 \\ 0 & \text{else} \end{cases}$$

Despite these horrible training results, it is still possible to choose a value of η so that your audiovisual fusion system has zero error. What value of η gives your classifier zero error?

Solution: Any value of η is OK for which $\{\eta \ln(2-2c) > 0\} \Leftrightarrow c < 0.5$, and this is true for any positive value of η.

22. (16 points) Good days and bad days follow each other with the following probabilities:

q_{t-1}	$p(q_t = G \vert q_{t-1} = \cdot)$	$p(q_t = B \vert q_{t-1} = \cdot)$
G	0.7	0.3
B	0.4	0.6

In winter in Champaign, the temperature on a good day is Gassian with mean $\mu_G = 50$, $\sigma_G = 20$. The temperature on a bad day is Gaussian with mean $\mu_B = 10$, $\sigma_G = 20$. A particular sequence of days has temperatures

$$\{x_1 = 10, x_2 = 20, x_3 = 30\}$$

What is the probability $p(X|q_1 = B)$, the probability of seeing this sequence of temperatures given that the first day was a bad day?

Solution:

$$p(X|q_1 = B) = \left(\frac{1}{50}\right)\left(\frac{1}{20}\right)\left(\frac{1}{20}\right)\left(\frac{6}{25}\right)((0.6)(0.35)(0.6) + (0.4)(0.13)(0.3)+$$
$$(0.6)(0.35)(0.4) + (0.4)(0.13)(0.7))$$

23. (25 points) Suppose that

$$\begin{aligned} a_{ij} &= p(q_t = j | q_{t-1} = i) \\ b_j(x_t) &= p(x_t | q_t = j) \\ g_t &= p(x_t | x_1, \ldots, x_{t-1}) \end{aligned}$$

And define the scaled forward algorithm to compute

$$\tilde{\alpha}_t(i) = p(q_t = i | x_1, \ldots, x_t) = \frac{p(x_t, q_t = i | x_1, \ldots, x_{t-1})}{g_t} = \frac{p(x_1, \ldots, x_t, q_t = i)}{g_1 g_2 \cdots g_t}$$

(a) Devise an algorithm to iteratively compute g_t and $\tilde{\alpha}_t(i)$. Fill in the right-hand side of each equation, using only the terms a_{jk}, $b_j(x_\tau)$, g_τ, and $\tilde{\alpha}_\tau(j)$ for $1 \le j \le N$, $1 \le k \le N$, $1 \le \tau \le t$.

1. **INITIALIZE:** $g_1 =$
2. **INITIALIZE:** $\tilde{\alpha}_1(i) =$
3. **ITERATE:** $g_t =$
4. **ITERATE:** $\tilde{\alpha}_t(i) =$
5. **TERMINATE:** $p(X) =$

Solution:

1. **INITIALIZE:** $g_1 = \sum_{j=1}^{N} \pi_j b_j(x_1)$
2. **INITIALIZE:** $\tilde{\alpha}_1(i) = \frac{\pi_i b_i(x_1)}{g_1}$
3. **ITERATE:** $g_t = \sum_{i=1}^{N} \sum_{j=1}^{N} \tilde{\alpha}_{t-1}(i) a_{ij} b_j(x_t)$
4. **ITERATE:** $\tilde{\alpha}_t(i) = \frac{\sum_{j=1}^{N} \tilde{\alpha}_{t-1}(i) a_{ij} b_j(x_t)}{g_t}$
5. **TERMINATE:** $p(X) = \prod_{t=1}^{T} g_t$

(b) Suppose $\beta_t(i) = p(x_{t+1}, \ldots, x_T | q_t = i)$. Then

$$\tilde{\alpha}_t(i)\beta_t(i) = p(f|g)$$

for some list of variables f, and some other list of variables g. Specify what variables should be included in each of these two lists.

Solution:

$$f = \{q_t = i, x_{t+1}, \ldots, x_T\}$$

$$g = \{x_1, \ldots, x_t\}$$

24. (20 points) The Maesters of the Citadel need to determine when winter starts. The temperature on day t is x_t. The state of day t is either $q_t = 0$ (Autumn) or $q_t = 1$ (Winter). Nobody really knows how cold this winter will be or how long it will last, but the Maesters have created an initial model $\Lambda = \{a_{ij}, b_j(x)\}$ where $a_{ij} \equiv p(q_t = j|q_{t-1} = i)$ and $b_j(x) \equiv p(x_t = x|q_t = j)$.

(a) Suppose we have a particular three day sequence of measurements, x_1, x_2, and x_3. Given that the preceding day was still autumn ($q_0 = 0$), we want to determine the joint probability that it continued to be autumn for days 1, 2, and 3, and that the three observed temperatures were measured. In other words, we want an estimate of

$$G_1 = p(q_1 = 0, x_1, q_2 = 0, x_2, q_3 = 0, x_3|q_0 = 0, \Lambda)$$

Find G_1 in terms of a_{ij} and $b_j(x_t)$, for whatever particular values of i, j, and t are most useful to you.

Solution:

$$G_1 = a_{00}^3 b_0(x_1) b_0(x_2) b_0(x_3)$$

(b) Suppose it is known that the preceding day was still autumn ($q_0 = 0$). Now, on day 1, the Maesters have determined that the temperature is x_1. Find the conditional probability, given this measurement, that it is still autumn, i.e., find

$$G_2 = p(q_1 = 0|x_1, q_0 = 0, \Lambda)$$

Find G_2 in terms of a_{ij} and $b_j(x_t)$, for whatever particular values of i, j, and t are most useful to you.

Solution:

$$G_2 = \frac{p(q_1 = 0, x_1|q_0 = 0, \Lambda)}{\sum_i p(q_1 = i, x_1|q_0 = 0, \Lambda)} = \frac{a_{00} b_0(x_1)}{a_{00} b_0(x_1) + a_{01} b_1(x_1)}$$

(c) The Maesters have collected a long series of measurements, $\{x_1, \ldots, x_T\}$ for T consecutive days. From these measurements, the Maesters have applied the forward-backward algorithm in order to calculate the following two quantities:

$$\alpha_t(i) \equiv p(x_1, \ldots, x_t, q_t = i|\Lambda), \quad \beta_t(i) \equiv p(x_{t+1}, \ldots, x_T|q_t = i, \Lambda)$$

Using these quantities, the Maesters wish to calculate the probability that Winter started on a particular day, $t = w$. That is, they wish to find

$$G_3 = p(q_{w-1} = 0, q_w = 1|x_1, \ldots, x_T, \Lambda)$$

Find G_3 in terms of $\alpha_t(i)$, $\beta_t(i)$, a_{ij} and $b_j(x_t)$, for whatever particular values of i, j, and t are most useful to you.

Solution:

$$G_3 = \frac{p(q_{w-1} = 0, q_w = 1, x_1, \ldots, x_T|\Lambda)}{\sum_i \sum_j p(q_{w-1} = i, q_w = j, x_1, \ldots, x_T|\Lambda)} = \frac{\alpha_{w-1}(0) a_{01} b_1(x_w) \beta_w(1)}{\sum_i \sum_j \alpha_{w-1}(i) a_{ij} b_j(x_w) \beta_w(j)}$$

25. (20 points) A bimodal HMM uses a common state sequence, $Q = [q_1, \ldots, q_T]$, to explain two different observation sequences $X = [\vec{x}_1, \ldots, \vec{x}_T]$ and $Y = [\vec{y}_1, \ldots, \vec{y}_T]$. The HMM is parameterized by

$$\begin{aligned} \pi_i &= p(q_1 = i) \\ a_{ij} &= p(q_t = j | q_{t-1} = i) \\ b_j(\vec{x}_t) &= p_X(\vec{x}_t | q_t = j) \\ c_j(\vec{y}_t) &= p_Y(\vec{y}_t | q_t = j) \end{aligned}$$

Define

$$\begin{aligned} \alpha_t(i) &= p(\vec{x}_1, \vec{y}_1, \ldots, \vec{x}_t, \vec{y}_t, q_t = i) \\ \beta_t(i) &= p(\vec{x}_{t+1}, \vec{y}_{t+1}, \ldots, \vec{x}_T, \vec{y}_T | q_t = i) \end{aligned}$$

(a) Specify initialization formulas for $\alpha_1(i)$ and $\beta_T(i)$ in terms of π_i, a_{ij}, $b_j(\vec{x}_t)$, and $c_j(\vec{x}_t)$.

Solution:

$$\alpha_1(i) = \pi_i b_i(\vec{x}_1) c_i(\vec{y}_1)$$
$$\beta_T(i) = 1$$

(b) Specify iteration formulas for $\alpha_t(i)$ and $\beta_t(i)$ in terms of π_i, a_{ij}, $b_j(\vec{x}_t)$, $c_j(\vec{x}_t)$, $\alpha_{t-1}(j)$, and $\beta_{t+1}(j)$.

Solution:

$$\alpha_t(i) = \sum_{j=1}^{N} \alpha_{t-1}(j) a_{ji} b_i(\vec{x}_t) c_i(\vec{y}_t)$$
$$\beta_t(i) = \sum_{j=1}^{N} \beta_{t+1}(j) a_{ij} b_j(\vec{x}_{t+1}) c_j(\vec{y}_{t+1})$$

26. (20 points) You are creating a recommender system that tries to recommend songs that will be considered to be similar to a given query. Each song is characterized by a two-dimensional vector $\vec{x}_k = [b_k, v_k]^T$ where b_k is the number of beats per minute, and v_k is the fraction of air-time during which there is a human voice. Your customer considers the following four songs to be similar:

$$[\vec{x}_1, \vec{x}_2, \vec{x}_3, \vec{x}_4] = \left[\begin{array}{cccc} 120 & 140 & 140 & 120 \\ 0.3 & 0.3 & 0.5 & 0.5 \end{array}\right]$$

You are given two more test data, $\vec{x}_5 = [b_5, v_5]^T$ and $\vec{x}_6 = [b_6, v_6]^T$, and you are asked whether or not $\vec{x}_5$ and $\vec{x}_6$ should be considered similar. Write formulas for the Mahalanobis distance between $\vec{x}_5$ and $\vec{x}_6$ under the following conditions:

(a) Estimate a diagonal data covariance matrix directly from the data, and use it to write the squared Mahalanobis distance $d^2_\Sigma(\vec{x}_5, \vec{x}_6)$.

Solution:

$$d^2_\Sigma(\vec{x}_5, \vec{x}_6) = \frac{(b_5 - b_6)^2}{100} + \frac{(v_5 - v_6)^2}{0.01}$$

(b) Estimate a diagonal data covariance matrix from the data, then regularize it using regularization parameter $\lambda = 0.01$ before using the result to write the squared Mahalanobis distance $d^2_\Sigma(\vec{x}_5, \vec{x}_6)$.

Solution:

$$d^2_\Sigma(\vec{x}_5, \vec{x}_6) = \frac{(b_5 - b_6)^2}{100.01} + \frac{(v_5 - v_6)^2}{0.02}$$